Count Unique Names in Excel

March 16, 2004
By Neil J. Rubenking

I have a list of some 900,000 names in an Excel worksheet. Most of the names have duplicates. How can I count the number of names without counting duplicate entries?

Brian Abbishaw

We'll assume that 900,000 is a bit of hyperbole, because an Excel spreadsheet cannot have more than 65,536 rows. For a few thousand names, there's a simple and elegant formula to count unique values—one that's worth knowing even if it won't handle tens of thousands.

First, highlight all the names and select Insert | Name | Define from the menu to give the range a name. I called it Names. Now enter this formula: =SUM(1/COUNTIF(Names,Names)). Now this is very important: Press Ctrl-Shift-Enter rather than only Enter; this will tell Excel to make it an array formula.

The second argument to COUNTIF is normally a single value, but we've passed an array. Excel substitutes each element of the array in turn and sums all the results. A unique name will return 1. When a name appears twice, each occurrence will add one-half to the sum—and one-half doubled is 1. No matter how many times a particular name appears, it will contribute exactly 1 to the total, so the result will be a count of unique names.

This formula breaks down when the data set is too big. The number of calculations Excel has to perform is proportional to the square of the number of items, so it gets slower and slower with more rows. In one test using a full 65,536 rows, it took almost 20 minutes. Clearly a different solution is needed.

Highlight the row containing the names and select PivotTable and PivotChart Report from the Data menu. Click on Next and then click on Finish. This will create an empty PivotTable in a new worksheet. Drag the Name field into the Drop Row Fields Here area, then drag it into the Drop Data Items Here area. Now you have a list of all the unique names, along with the number of times each name occurs. To get a count of unique names, enter this formula in a cell that's not in column A: =COUNTA(A:A)-3.

COUNTA returns the number of nonblank cells in the column. The two PivotTable headers and the grand total line are not counted, so we subtract 3. One warning: PivotTables are not recalculated automatically. If the list of names changes, click on the Refresh Data button (the red exclamation point) in the PivotTable toolbar.